Display Psychophysics and Observer Performance

Image Perception and Image Evaluation

- Human visual system
  - Luminance response
  - Spectral response
  - Spatial resolution response
  - Temporal response (flicker)
  - Noise and clutter
- Interaction of imaging system with visual system
  - Spatial resolution
  - Contrast resolution
- Digital representation of medical images
- Evaluation of image quality for perceptive tasks
  - Subjective quality of image appearance
  - Objective quality of human performance

Image Perception

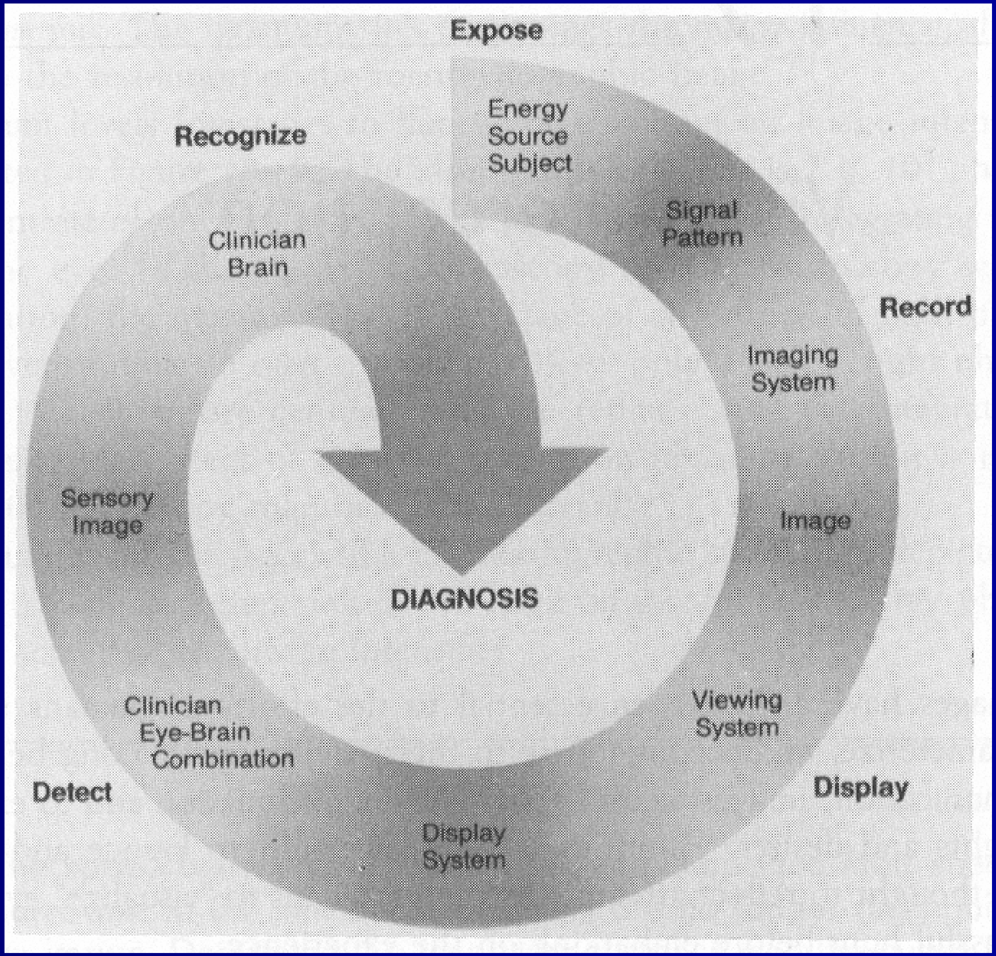
"Medical imaging is partway between science and art. Both activities are attempts to transmit to the eye and brain of observers some more or less abstract *impression* of an object of interest and, in so doing, somehow to influence their state of mind."

"...the ultimate goal can only be reached through the operation of the particular properties of two sequential channels: the visual faculties of the observers and their mental processes."

C.R. Hill, "Perception and Interpretation of Images", in: The Physics of Medical Imaging, S. Webb, ed., Hilger, 1990.

Image Perception

Medical images are acquired in order to provide information to a physician that aids in the formation of a diagnosis.



Human visual system – luminance

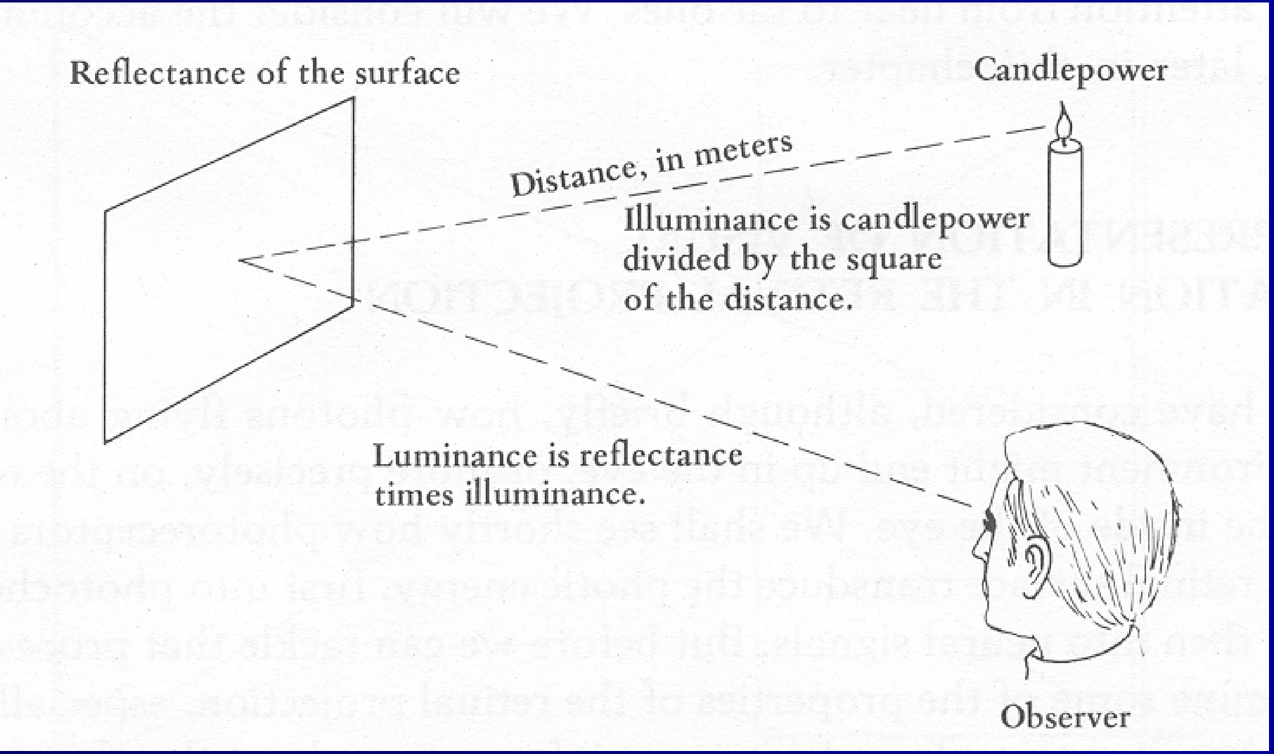
Light is described differently in *radiometry* (physical units) and *photometry* (psychophysical units).

Radiometry (Physical units)	
Definition	**SI unit**
Radiant energy	joule
Radiant density	joules per cubic meter
Radiant flux or power	watt
Radiant emittance	watts per square meter
Irradiance	watts per square meter
Radiant intensity	watts per steradian
Radiance	watts per steradian and square meter

Photometry (Psychophysical units)	
Definition	**SI unit**
Luminous energy	lumen second (talbot)
Luminous density	lumen seconds per cubic meter (talbot/m^3)
Luminous flux or power	lumen
Luminous emittance	lumens per square meter (lux)
Illuminance	lumens per square meter (lux)
Luminous intensity	candela (lumens per steradian)
Luminance	lumens per steradian and square meter (nit)

Human visual system – illumination & luminance

Luminance is the quantity of light reaching the visual system. The eye can respond to a range of intensities differing by a factor of 10^5.



Human visual system – luminance

Units of illumination

1 metre-candle	= 1 lux	= 1 lumen per sq. m.
1 phot	= 10,000 lux	
1 milliphot	= 10 lux	
1 foot-candle	= 10·764 lux	= 1 lumen per sq. ft.

Units of luminance

1 candela per sq. cm.	= 10,000	candelas per sq. m.
1 stilb	= 10,000	,, ,, ,,
1 nit	= 1	,, ,, ,,
1 candela per sq. ft.	= 10·764	,, ,, ,,
1 candela per sq. in.	= 1550	,, ,, ,,
1 equivalent phot	= 3183	,, ,, ,,
1 lambert	= 3183	,, ,, ,,
1 millilambert	= 3·183	,, ,, ,,
1 equivalent lux	= 0·3183	,, ,, ,,
1 blondel	= 0·3183	,, ,, ,,
1 apostilb	= 0·3183	,, ,, ,,
1 equivalent foot-candle	= 3·426	,, ,, ,,
1 foot-lambert	= 3·426	,, ,, ,,

Human visual system – luminance

Luminance of a light-coloured object under various illuminations

Direct sunlight - - -	25,000	cd./sq. m.
Daylight out-of-doors - -	10^3 to 10^4	,,
Good interior lighting - -	100 to 1000	,,
Moderate interior lighting - -	10 to 100	,,
Feeble interior lighting - -	1 to 10	,,
Outdoor lighting by night in a town	0·1 to 1	,,
Night vision - - -	10^{-4} to 0·1	,,

Film viewbox	~1500	cd/m2
Highlights of bright CRT	1000	cd/m2
Dark area of low-level CRT	0.1	cd/m2

Human visual system – luminance

Retinal illumination often considers the size of the pupil as well: a *troland* is 1 candela/m^2 * 1 mm^2

Pupil diameters and areas at various luminances

L : luminance d : pupil diameter S : pupil area
S_e : effective pupil area

L (candelas/ sq. m.)	d (mm.)	S (mm.^{2})	S_e (mm.^{2})	$L\,.\,S$ (trolands)	$L\,.\,S_e$ (effective trolands)
1	5·00	19·7	15·0	19·7	15
2	4·64	16·9	13·4	34	27
5	4·18	13·7	11·3	69	57
10	3·86	11·7	9·96	117	100
20	3·57	10·0	8·72		174
50	3·23	8·17	7·30		365
1×10^2	3·01	7·12	6·46		646
2×10^2	2·82	6·24	5·73		$1{\cdot}15\times10^3$
5×10^2	2·62	5·39	5·01		$2{\cdot}5\times10^3$
1×10^3	2·50	4·91	4·59		$4{\cdot}6\times10^3$
2×10^3	2·40	4·52	4·25		$8{\cdot}5\times10^3$
5×10^3	2·30	4·15	3·92		$1{\cdot}96\times10^4$

This table seems wrong, doesn't it?

The *Stiles-Crawford* effect states that the pupil acts as though its effective area is smaller, because off-center illumination seems dimmer than centered.

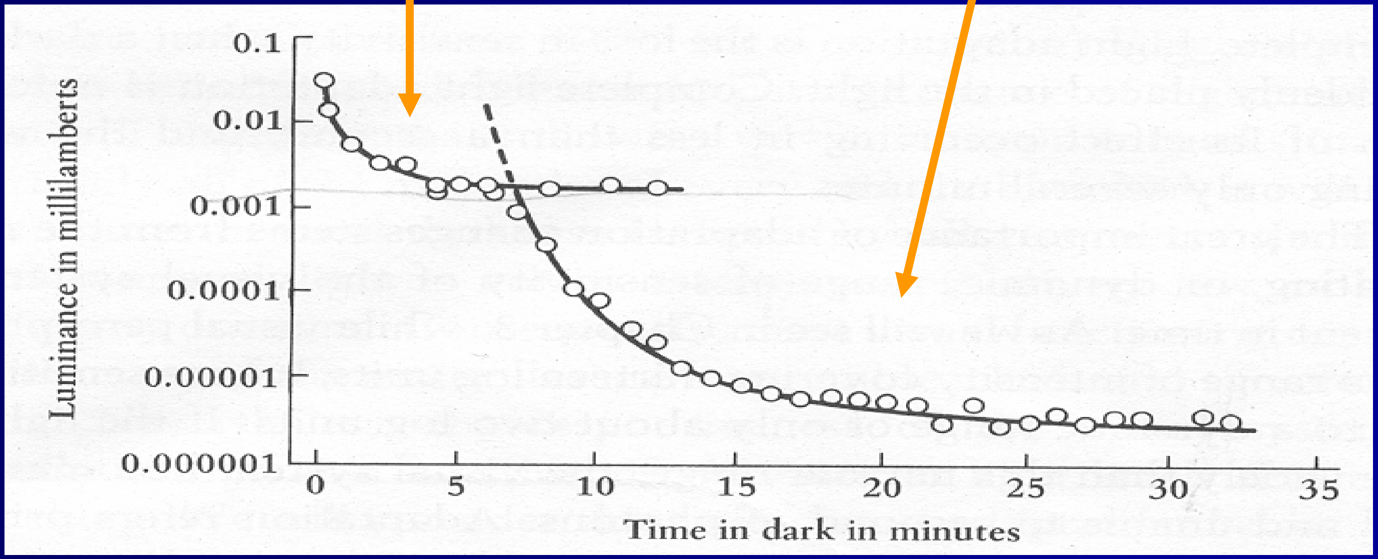
Human visual system – luminance response

Photopic (good lighting) vs. scotopic (dark-adapted) vision. Threshold luminance as a function of time of dark adaptation.

Threshold of illumination is approximately equivalent to a candle flame at 30 miles on a dark clear night.

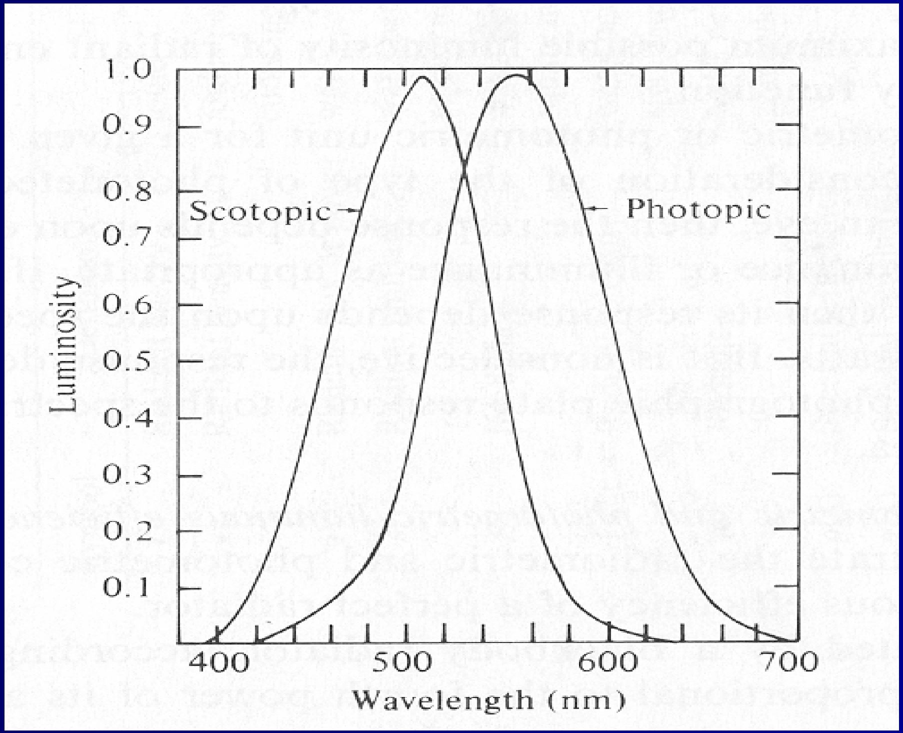
Cones

Rods



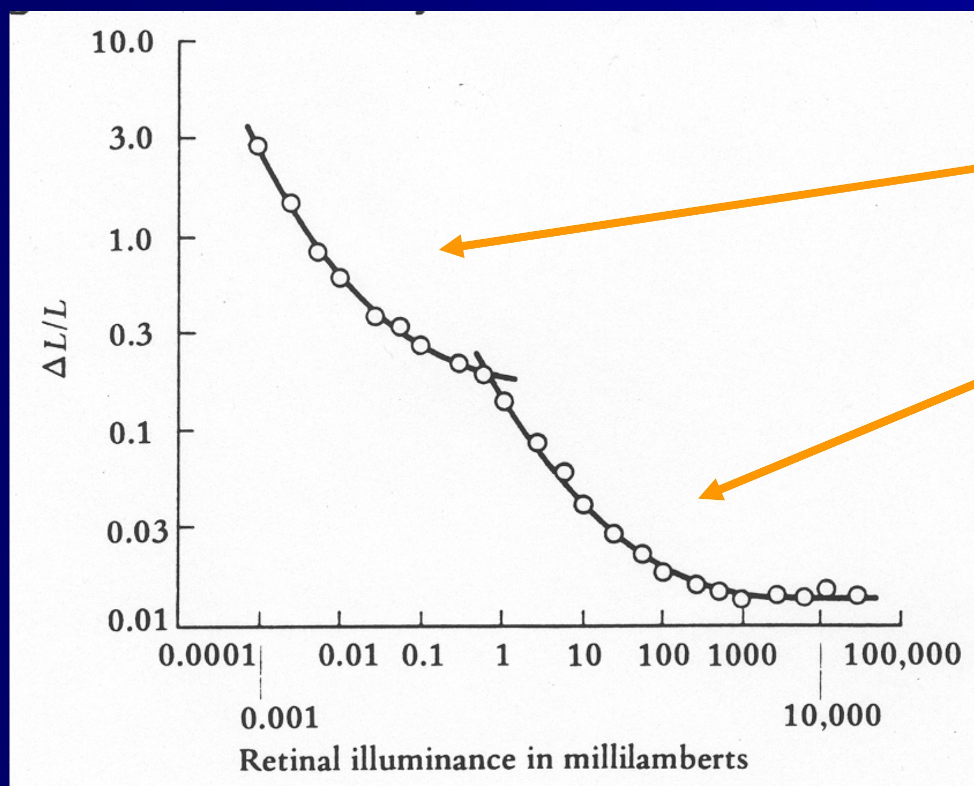
Human visual system – luminance response

Photopic vs. scotopic vision: the Purkinje shift.



Human visual system – luminance response

Weber ratio - threshold of ΔL that just permits areas of L and L +ΔL to be distinguished.



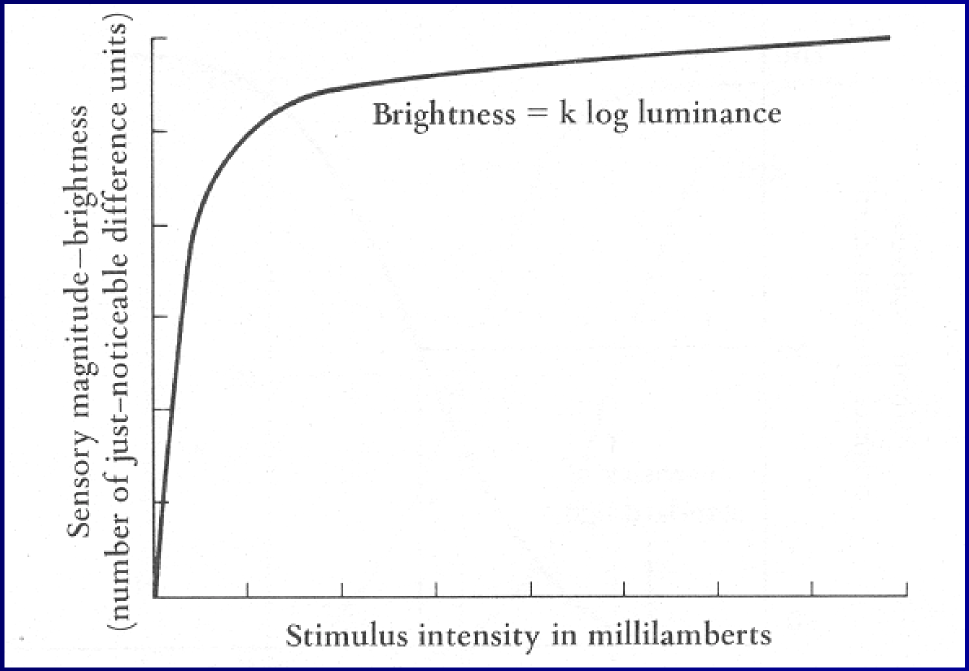
rods

cones

Human visual system – luminance response

At ordinary levels of luminance, effective range is about 1000:1 in luminance, vs. 10 billion:1 over full range.

Number of just-noticeable-differences (jnds) in luminance varies with average luminance.



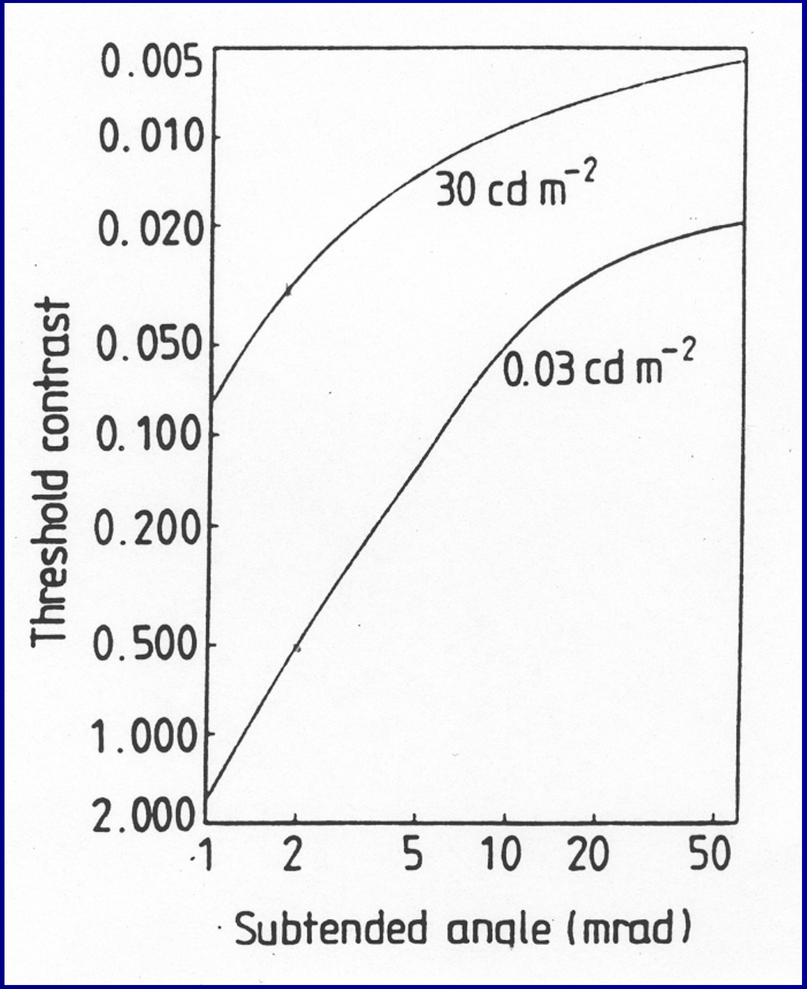
Human visual system – luminance response

Angular size of object also affects detectability

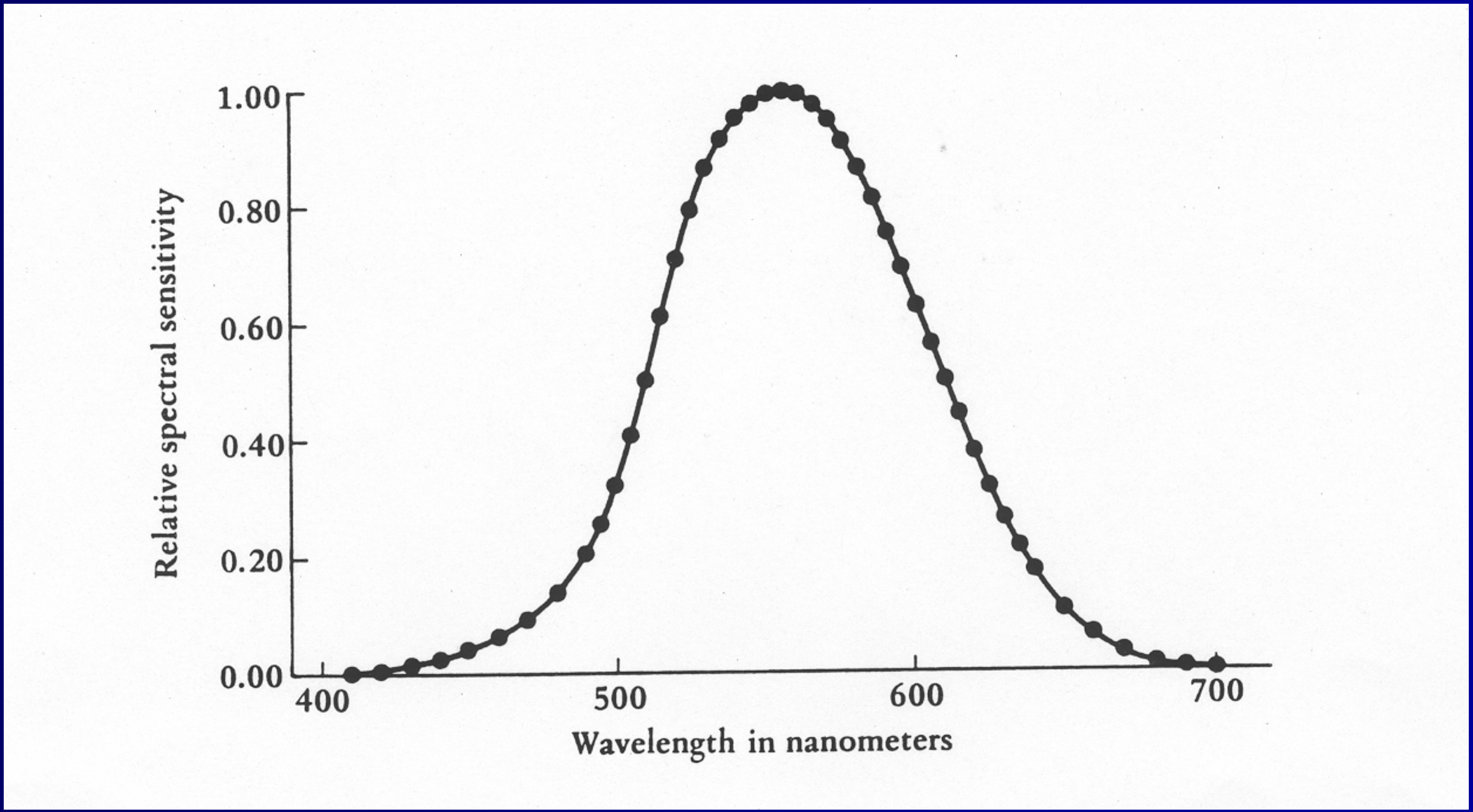
Threshold contrast of an object = ΔL / L

Units:

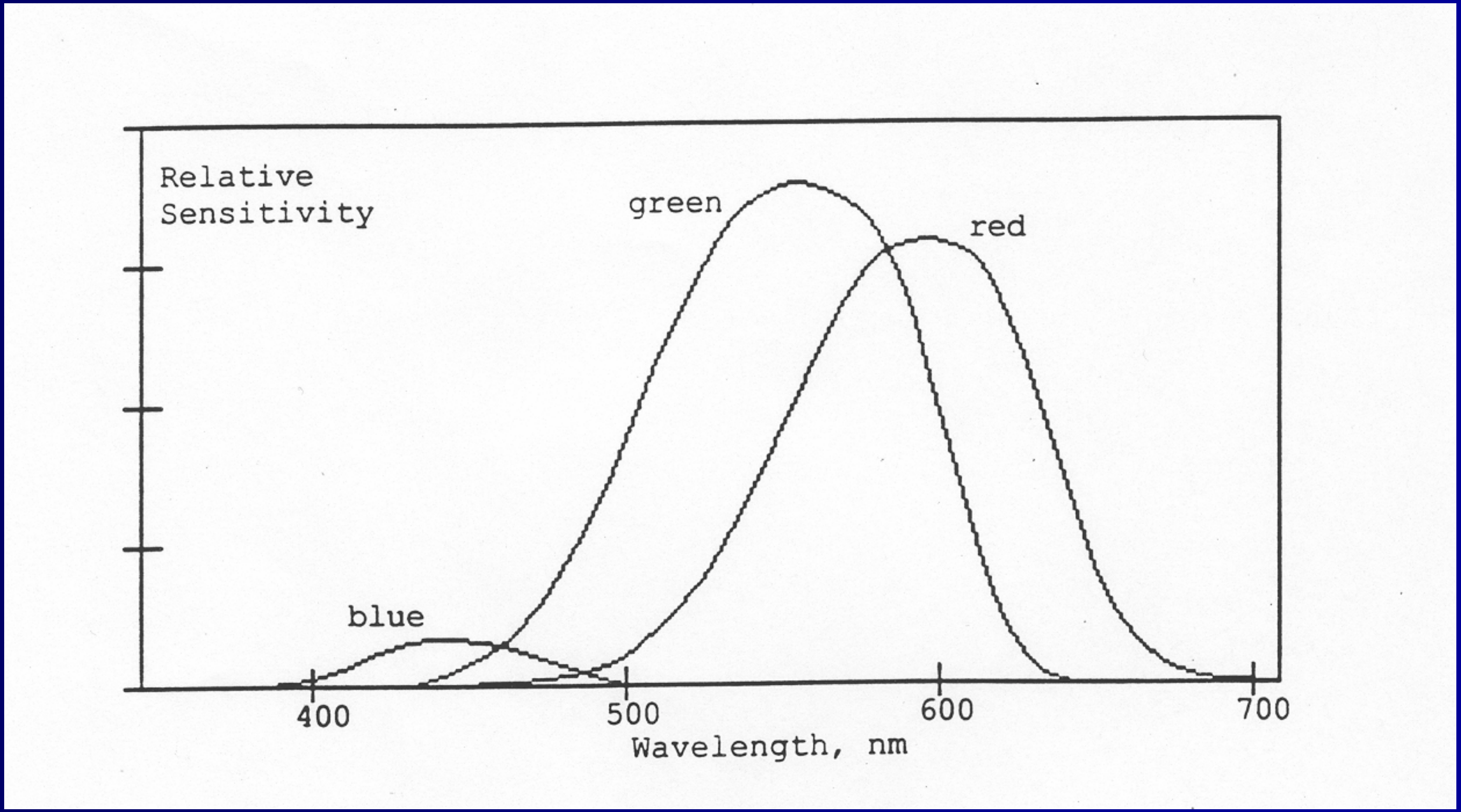
1 milliradian = .057 degrees



Human visual system – spectral response



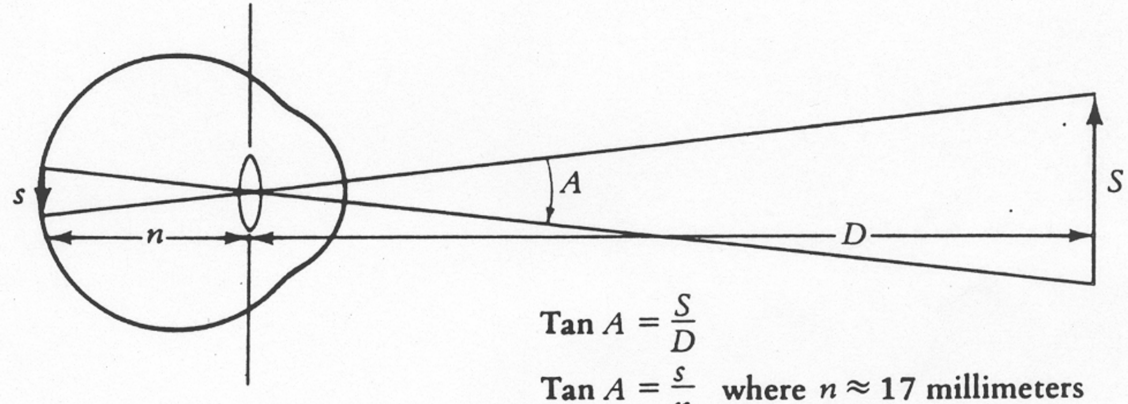
Human visual system – spectral response



Human visual system – spatial resolution

The fovea has the highest spatial resolution capability.

- Limited by cone size & packing structure to 60 cycles/degree under best conditions.
- The fovea subtends about 2 degrees from the plane of the lens.



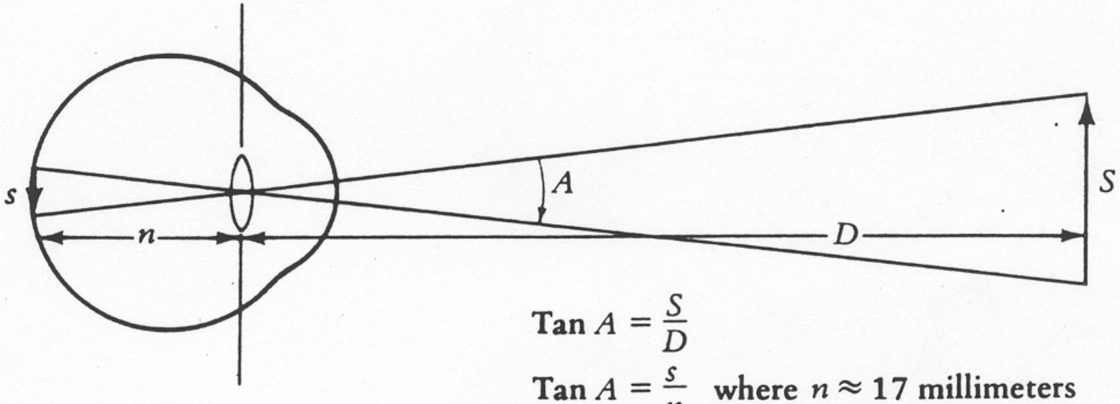
$$\text{Tan } A = \frac{S}{D}$$

$$\text{Tan } A = \frac{s}{n} \quad \text{where } n \approx 17 \text{ millimeters}$$

$$s = \frac{17S}{D} \quad \text{in millimeters}$$

Human visual system – spatial resolution

- At a viewing distance of 16" (40.6 cm), an area of ~ 0.5 cm in diameter is imaged onto the fovea.
- For an image with spatial resolution of 10 line-pairs/mm, 50 cycles/degree are imaged onto the fovea.



$$\text{Tan } A = \frac{S}{D}$$

$$\text{Tan } A = \frac{s}{n} \quad \text{where } n \approx 17 \text{ millimeters}$$

$$s = \frac{17S}{D} \quad \text{in millimeters}$$

Human visual system – spatial resolution

- Acuity
  - 20/20 vision (acuity = 1.0); can see a line subtending 1 minute of arc at 20 feet (1.8 mm long)
  - 20/100 vision (acuity = 0.2); line would subtend 1 minute at 100 ft, or 5 minutes at 20 ft.
- Thumbnail at arm's length subtends about 1.5 – 2 degrees.

Human visual system – spatial resolution

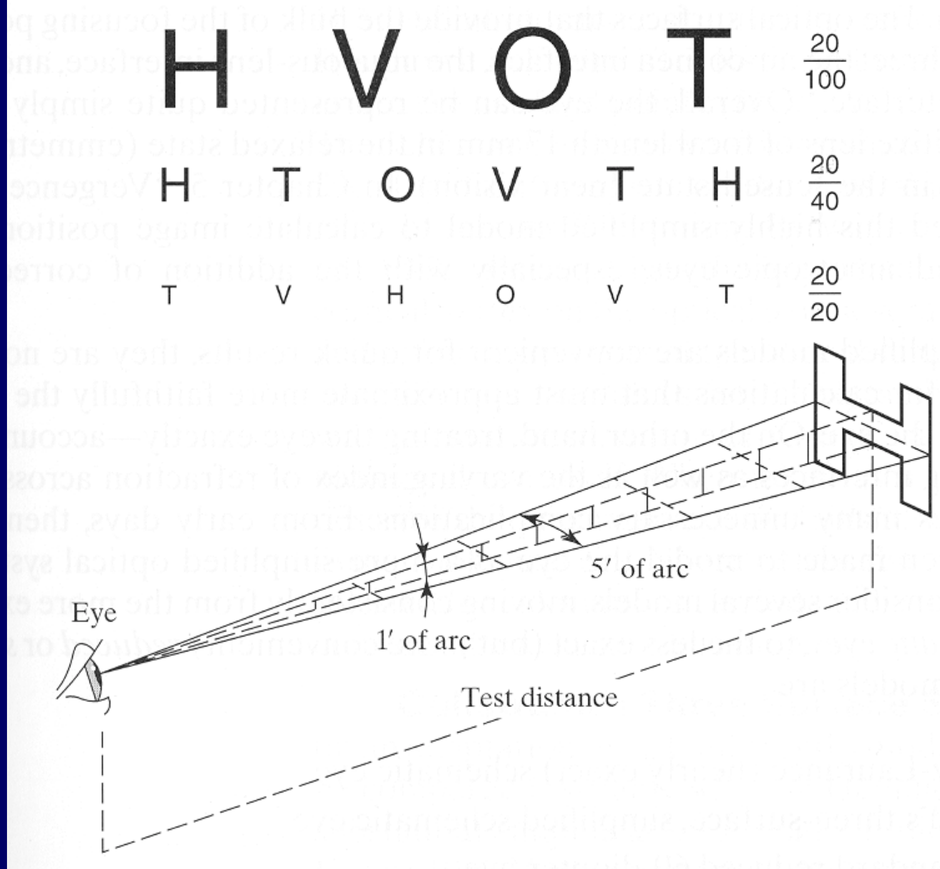
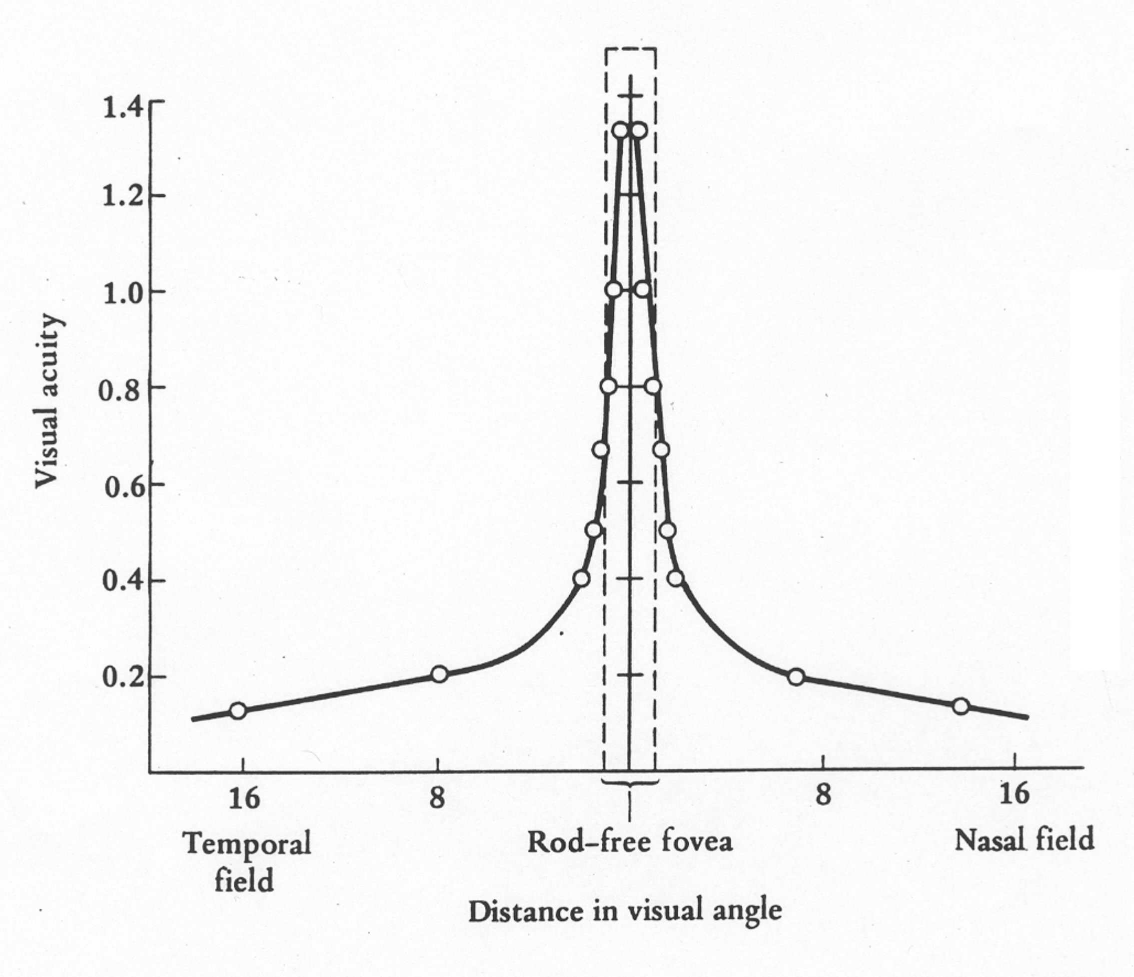


Figure 10-2 Construction of a Snellen eye-chart letter H to measure visual acuity. Notice the measure of the 5′ angle for the entire letter and the measure of the 1′ angle for the width of any member. The top portion of the figure shows a section of an eye chart (reduced) containing the letter H and several other letters, with 20/20, 20/40, and 20/100 rows of letters identified.

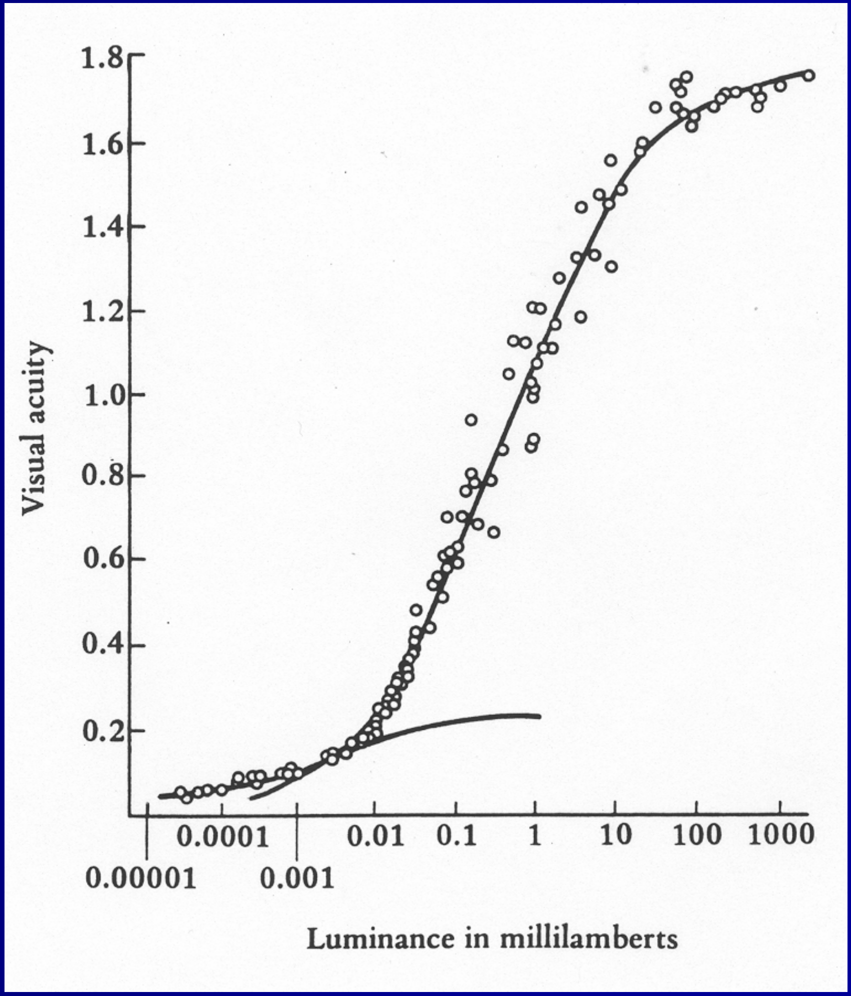
Human visual system – spatial resolution

Foveal vs. peripheral acuity



Human visual system – spatial resolution

Acuity is also a function of luminance.



Human visual system – spatial resolution

To achieve foveal scanning of a chest radiograph (36 x 43 cm) in .5 x .5 cm swatches (at viewing distance of 16 inches) would require 6200 foveal fixations. At 3 fixations per second that would take 34 minutes!

One model for how radiologists can read in 60 seconds:

- global analysis (symmetry, big abnormalities)
- checking fixations (verify their first impressions)
- discovery scanning (search relevant areas)
- reflective scanning (review)
- decision

Human visual system – spatial resolution

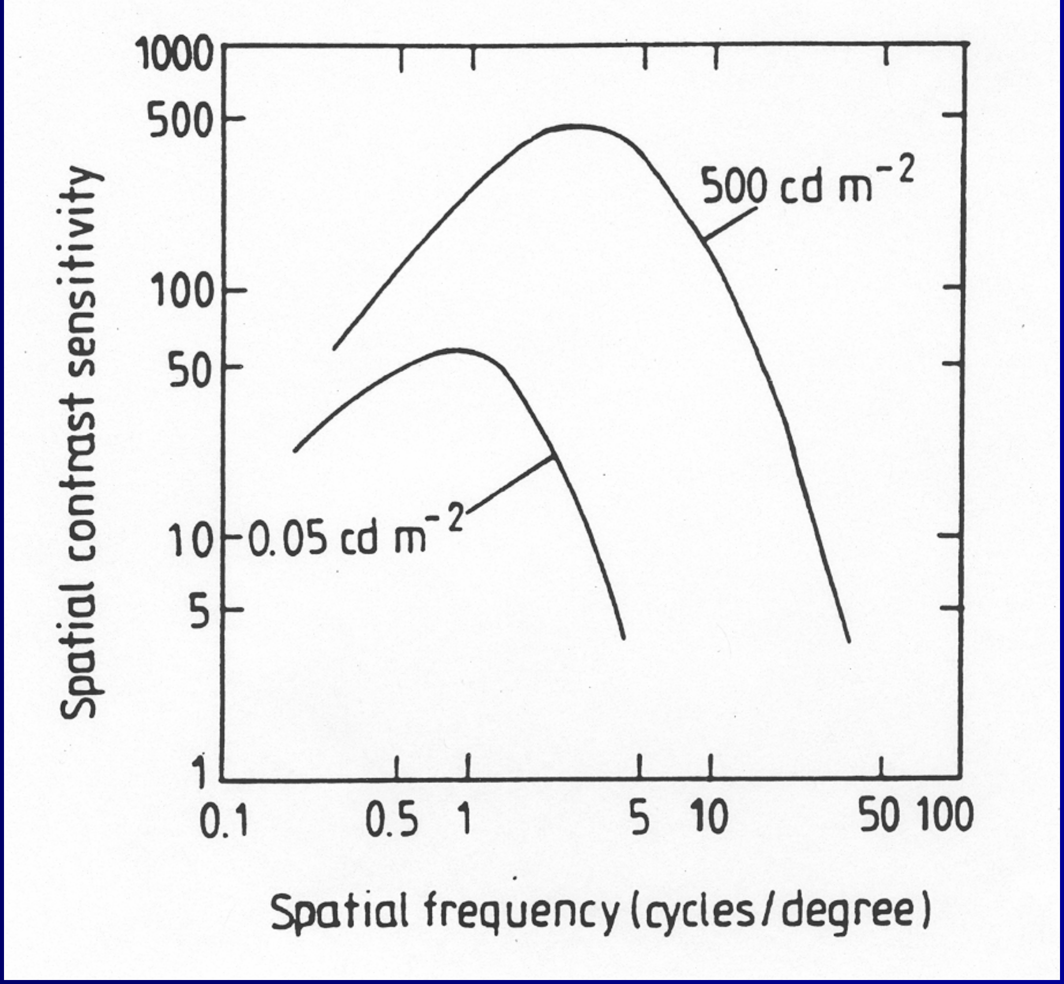
Affected by the luminance of the scene

– Film viewbox	~1500 cd/m^2
– Highlights of bright CRT	1000 cd/m^2
– Dark area of low-level CRT	0.1 cd/m^2
– Comfortable reading	30 cd/m^2

Typical CRT has a max/min luminance ratio of 100:1 or less. Example: 2 > 207 cd/m^2 (measured with a test pattern).

X-ray film typically has a ratio of ~1000:1 (optical density of 0.15 to 3.2 or, transmission of 0.06% to 70% of light)

Human visual system – spatial resolution



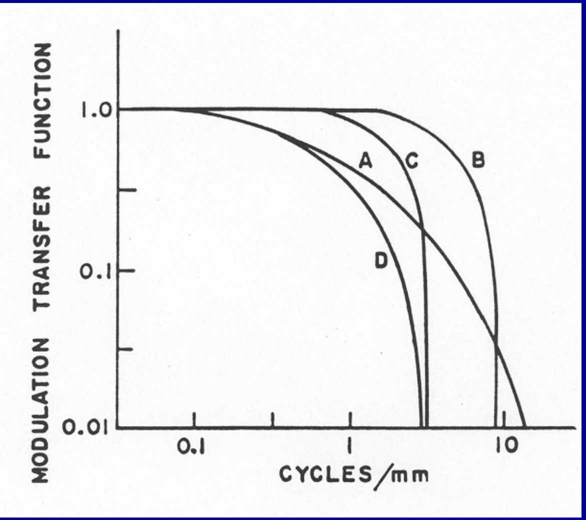
Imaging systems – spatial resolution

Capture resolution -- signal pattern is recorded by an imaging system with a characteristic modulation transfer function.

Displayed resolution -- the recorded image is often output for human interpretation by a display device that has its own MTF.

Composite MTF is the product of the MTFs of all components of the system.

Presentation of spatial frequency information to the user should accommodate the imaging characteristics of the visual system.



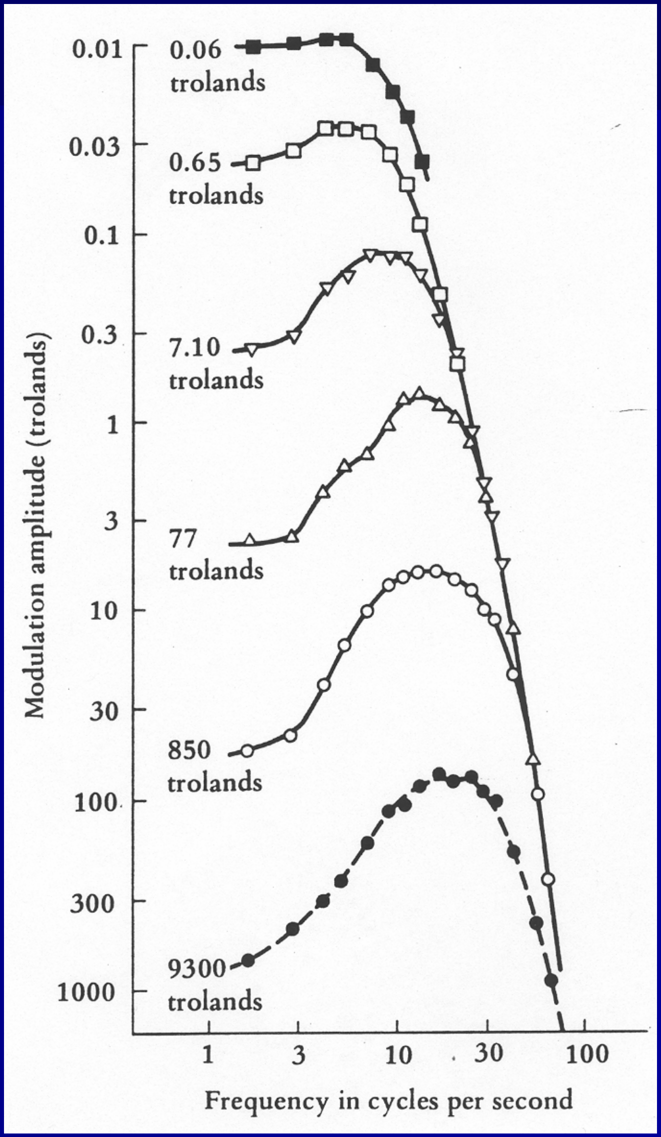
Human visual system –flicker

Luminance fluctuates:

$L(t) = L + \Delta L \cos(\pi f t)$

Humans are unable to detect a variation in luminance if ΔL is small enough or the temporal frequency (f) of the variation is high enough.

Flicker fusion rate depends on average L – six levels are shown.



Human visual system – noise and clutter

Detection of an object against a background when both are noisy -- the signal to noise ratio needs to be ~ 2 to 5 for good probability of detection.

$$\mathrm{SNR} = (\bar{S} - \bar{B}) \,/\, \mathrm{SQRT}(\sigma^2{}_S + \sigma^2{}_B)$$

$$= (C - \bar{B}) \,/\, \mathrm{SQRT}(\sigma^2{}_S + \sigma^2{}_B)$$

Structured noise or "clutter" also affects the ability to detect objects.

Human visual system – noise and clutter

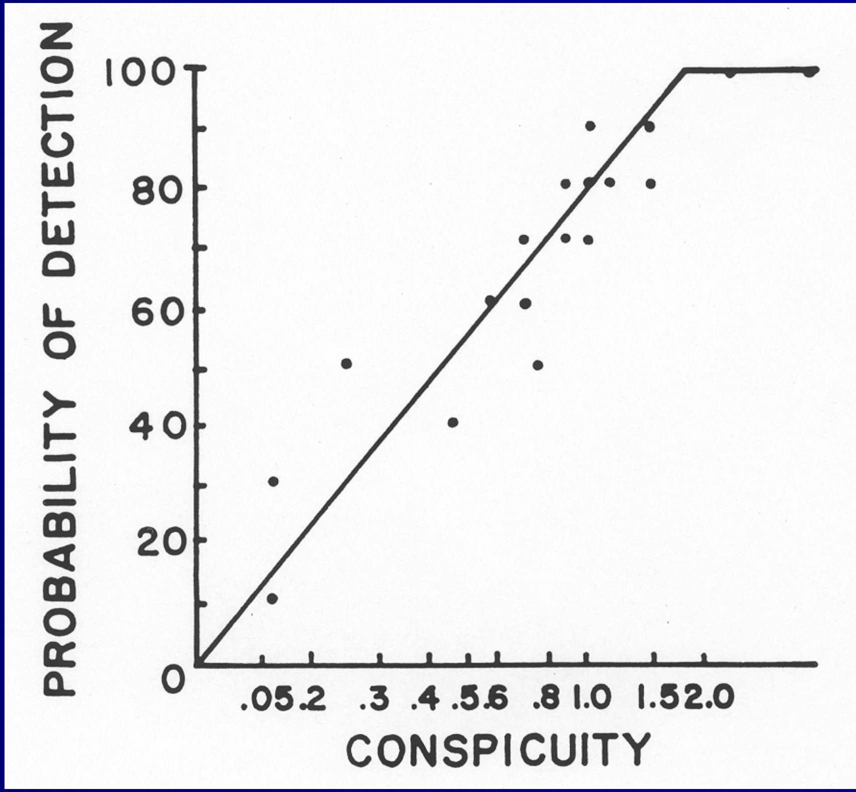
Figure 8. An example of camouflage. Grouping together of many seemingly distinct features in the image and ignoring apparent similarities between others allows the eye to find the real structures present.

Human visual system – noise and clutter

Conspicuity (contrast/complexity) attempts to describe the effect of clutter – structured noise – on detection.

A simple measure of complexity is fluctuation of density around a target.

Consider subtraction angiography or dual-energy subtraction as methods to reduce clutter.



Imaging systems – spatial resolution

Full greyscale range

a) 256 x 256 pixels

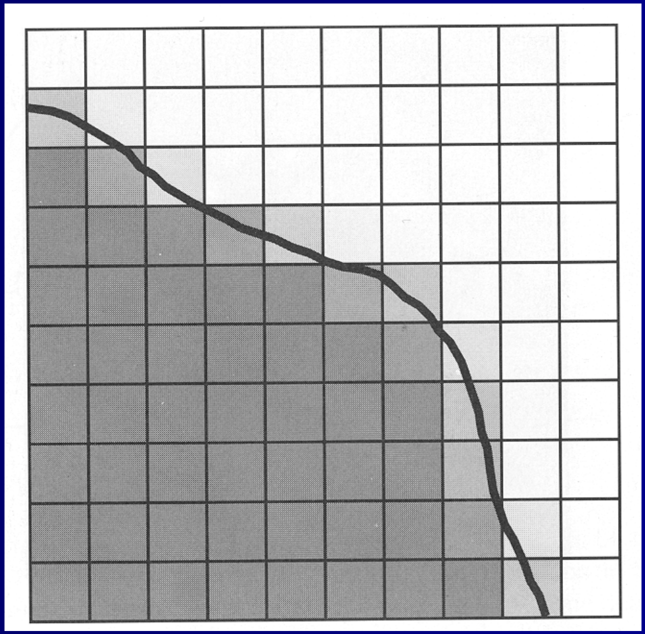
b) 128 x 128

c) 64 x 64

d) 32 x 32

Imaging systems – contrast resolution

Quantizing the signal strength of an image into a limited number of display output levels affects the representation of boundaries.



Imaging systems – contrast resolution

Too few grey levels used in presentation may result in artificial contours or jagged edges.

"Too many" grey levels may mean that effort is being wasted in representing noise. Or, if the visual system is unable to distinguish output levels, they convey no information.

Imaging systems – contrast resolution

a) 32 grey levels
b) 16 grey levels
c) 8 grey levels
d) 4 grey levels

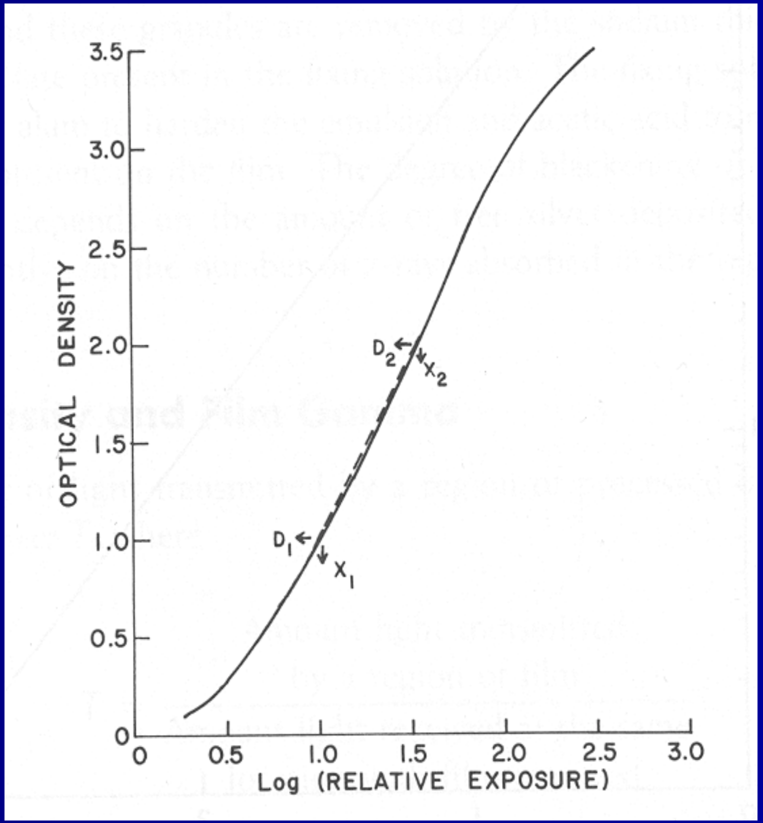
Digital representation of medical images

Procedure	Abnormality	Display	Pixel size
Chest 36 x 43 cm	nodules interstitial nodules various	film film video video	.2 – 1 mm < .1 - .45 mm .4 - .5 mm < .2 - <1.3 mm
Mammography 20 x 25 cm	microcalcifications	film	< .1 - .2 mm
Skeletal	various various	film video	.04 mm < .25 mm
Gastrointestinal	various	film	.4 mm
Genitourinary	various	film	.2 mm

Digital representation of medical images

Hurter-Driffield characteristic curve for an x-ray film.

The average gradient for the film is 1.9 over the density range 1.0 to 2.0.



Digital representation of medical images

Contrast resolution -- digitizing from film

grey level should represent $< 2 * \sigma(D)$

where D = optical density

film:

8 bits for linear part of H & D curve +

2 bits for nonlinear portions.

1024 grey levels total.

Digital representation of medical images

Procedure	pixel size	pixel matrix	bits	Mb
Chest	.2 mm .1 mm	1800 x 2150 3600 x 4300	10 10	7.7 31
Mammo	.1 mm	2000 x 2500	10	10
Skeletal	.04 mm	5000 x 6250	10	63

Digital workstations for display of medical images

For monitors, consider:

- Resolution
  - number of lines
  - pixels per line
- Contrast ratio
  - range (max / min for full-field)
  - detail (max / min for small areas)
- Phosphor
  - spectral band; particularly for side-by-side monitors
  - brightness
  - decay time
- Physical size and orientation

Digital workstations for display of medical images

Design:

- How many monitors are needed depends on how many images are needed simultaneously
- How will images be fetched & arranged?
- How fast can an image be fetched?
- What image manipulations are useful?
- What non-image information is useful?

Evaluation of the quality of images or workstations

Completely objective metrics, such as MTF and contrast ratios are often not good predictors of the ability of human to perform a task.

Subjective acceptance of image appearance or workstation:

- judge image quality by its ability to render anatomical details
- judge effectiveness by ease of use (speed, frequency, opinions)

But… subjective metrics are not very good predictors of human performance either.

Evaluation of the quality of images or workstations

Objective assessment of quality by measuring human performance on a specific task is necessary to determine the effect of the system on diagnostic ability.

But… objective assessments made under experimental conditions don't always reflect system performance in clinical use, and it can be difficult to generalize results to new situations.

Evaluation of Observer Performance

- Define the diagnostic task
- Choose a gold standard
- Choose a measure of performance
- Assemble a set of images
- Assemble a set of observers
- Control unwanted sources of variation
- Run the observer study
- Perform statistical analysis of the results

Evaluation of Observer Performance – task

- Diagnosis of a specific disease
  - – e.g., lung tumor
  - – pneumothorax
- Observer classifies image as:
  - – Normal; no disease present
  - – Abnormal; disease present

Evaluation of Observer Performance – gold standard

Comparison of methods -- choose one as the standard for judging the other

e.g., Conventional screen/film vs. Computed Radiography system

or, Aided vs. unaided nodule detection

Observer Performance – ROC curves

Classification ability measured from

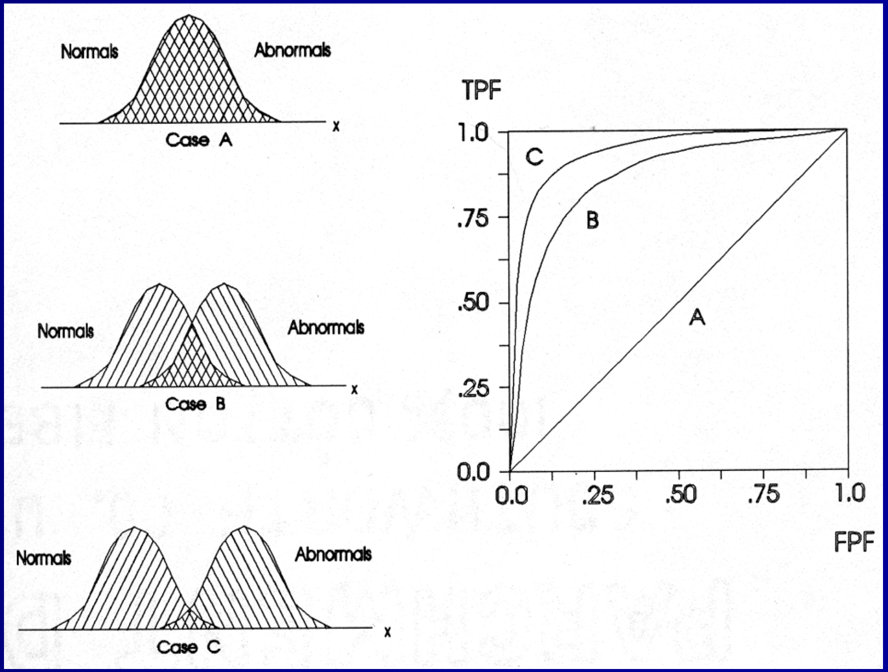
- standard ROC curve
- Localization ROC curve (LROC)
- Free response ROC curve (FROC)

Observer Performance – ROC curves

Area under standard ROC curve = A_Z is usually calculated from a fitting program that assumes bi-normal distributions.

Calculating the area under a piecewise linear representation of the curve is independent of distribution, but will underestimate true area.

Partial areas are also reported, e.g. area up to FPF = 0.20

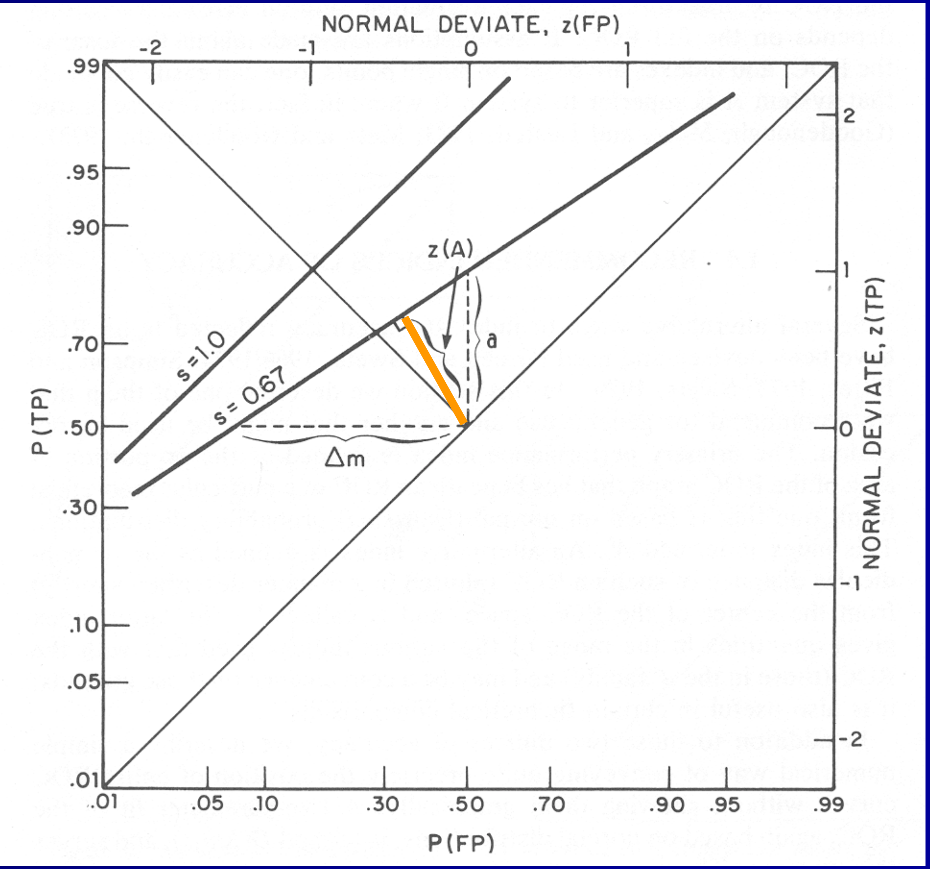


Observer Performance – ROC curves

A bi-normal ROC curve is sometimes reported as $d_a = z(A) * SQRT(2)$,

where z(A) is the perpendicular distance from center of graph to the ROC curve (orange line).

d_a ranges from 0 to 4 or 5



Observer Performance – 2AFC

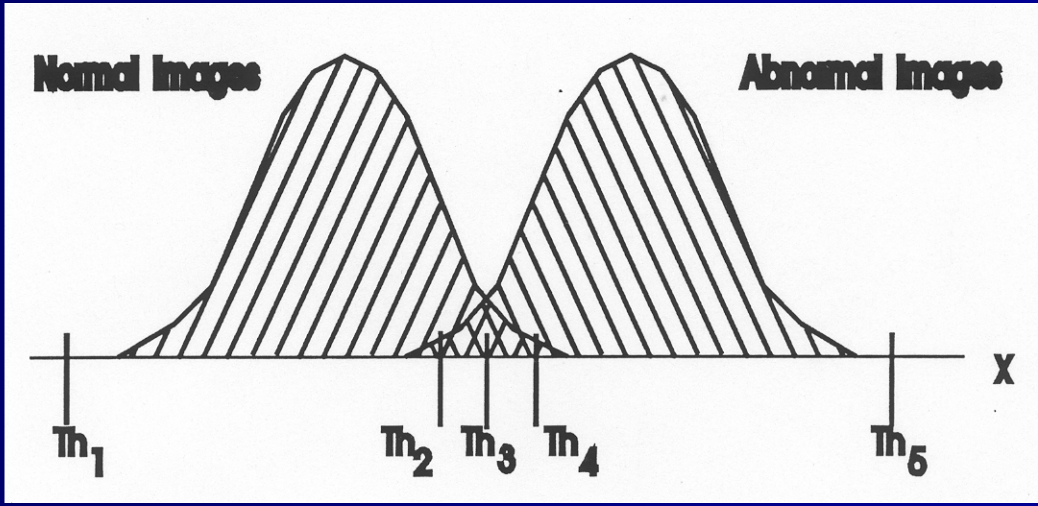
Area under standard ROC curve = A_Z is equal to the probability of a correct response in a 2-alternative forced-choice (2AFC) experiment.

In 2AFC, a trial consists of presenting one sample from each population, e.g. normal and abnormal images. The observer knows this and simply selects which of the pair is most likely to be the abnormal one.

2AFC is attractive due to its simplicity, but is limited because the entire ROC curve is not obtained, and it usually requires more trials to attain the same confidence interval about the estimated area.

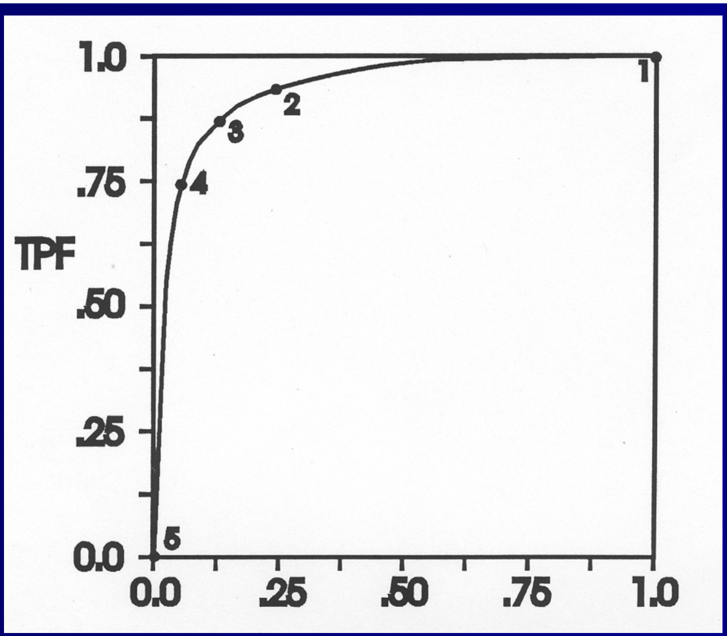
Observer Performance – ROC curves

An ROC curve is estimated by adjusting the decision threshold and recording the TPF and FPF for each threshold.



Observer Performance – ROC curves

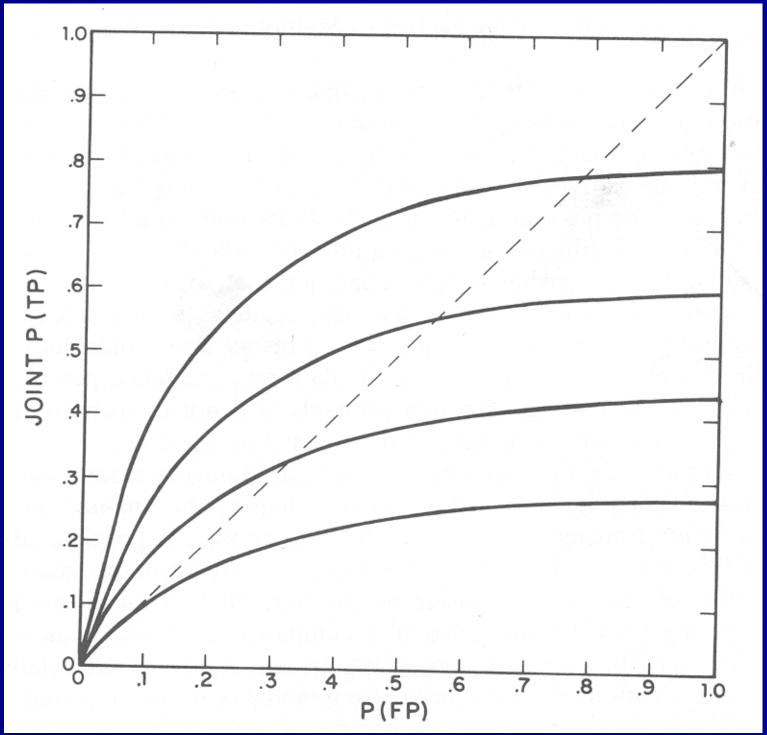
Thresh:	1	2	3	4	5
TPF	1.0	.95	.88	.75	0.0
FPF	1.0	.27	.15	.08	0.0



Observer Performance – LROC curves

Localization-response curves require both correct detection and correct localization of an abnormality to consider it a “true-positive” response.

There are 1 or 0 targets per image; location responses may be a forced choice between *m* locations.



Observer Performance – FROC curves

Free-response experiments allow reporting multiple targets in an image, with abnormality ratings for each.

Curves map true-positive detections as a function of the average number of false-positive detections per image.

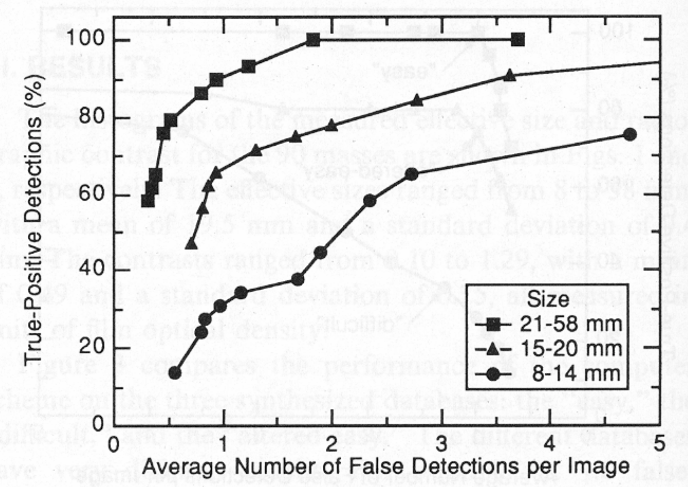


FIG. 5. The performance of the computer detection scheme when the database is divided into 3 groups of approximately equal number based on the effective size of the masses.

Evaluation of Observer Performance – image set

Truth of diagnosis is required:

- simulation
- biopsy
- complementary examinations
- repeat examinations
- consensus of experts' opinions

Variety of difficulty is required:

- some abnormalities that are easily detected
- some that are difficult to detect
- 70% - 80% correct classification is desirable

Images should be representative of larger population

Evaluation of Observer Performance – observers

- Must be familiar with the imaging techniques under examination
- Must have sufficient experience with the diagnostic task
- Experience should be current

Evaluation of Observer Performance – unwanted variation

Devices:
- Stabilize operating parameters
- X-ray equipment, film stock & developers,
- Computed Radiography system, video monitors

Working conditions:
- Ambient lighting, noise, and temperature

Images:
- Match cases, either directly or by complexity

Observers:
- Match directly, or by experience

Evaluation of Observer Performance – observer study

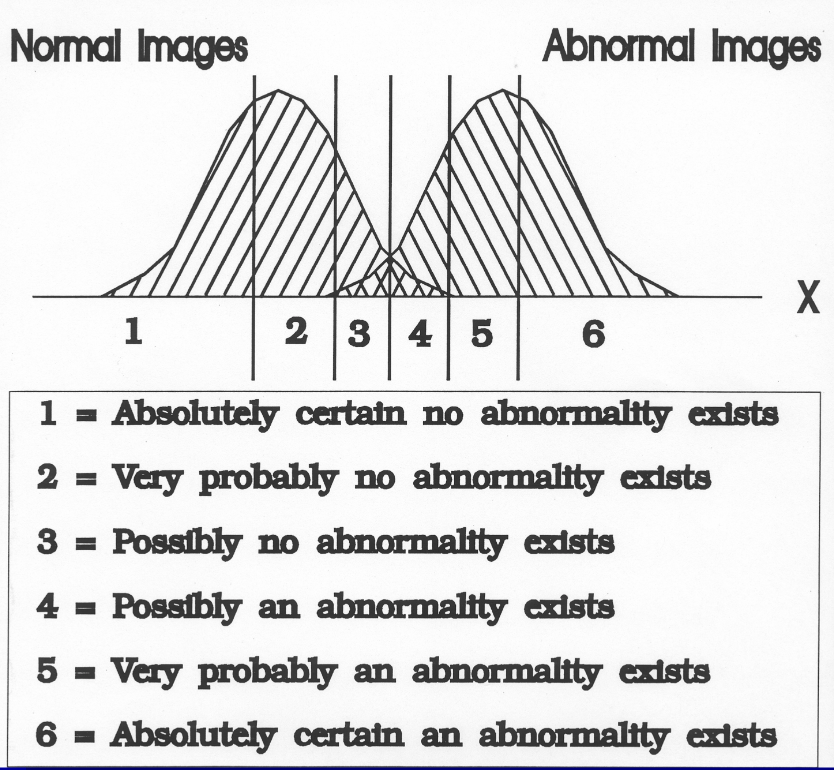
- Instruct the observers in the experimental diagnostic procedure e.g., confidence ratings for ROC analysis

- Collect observer responses without unintentional feedback by the researcher

- Avoid unduly fatiguing the observers

Evaluation of Observer Performance – analysis

Translating category responses into TPF and FPF.



cat	1	2	3	4	5	6
A	1	2	3	7	15	30
N	20	10	5	3	2	1

OP	1	2	3	4	5
TPF	.98	.95	.90	.78	.52
FPF	.51	.27	.15	.07	.02

Operating points at 0,0 and 1,1 assumed.

Evaluation of Observer Performance – analysis

ROC analysis for each observer yields:

Θ - estimated area under the ROC curve

σ_{Θ}^{2} - estimated variance in the estimated area

- estimated ROC curve

ROC analysis for the group of observers yields:

$\overline{\Theta}$ - estimated average area for the group

$\sigma_{\overline{\Theta}}^{2}$ - estimated variance in the average area

- estimated group ROC curve

Evaluation of Observer Performance – analysis

Hypothesis testing:

Is the difference $\bar{\Theta}_1 - \bar{\Theta}_2$ significant or due to random fluctuations in the data

Critical ratio $$z = \frac{\left(\bar{\Theta}_1 - \bar{\Theta}_2\right)}{S.E.\left(\bar{\Theta}_1 - \bar{\Theta}_2\right)}$$

Paired-difference t-test $$t = \frac{\bar{d}\,\sqrt{N}}{S.E.\left(d\right)}$$

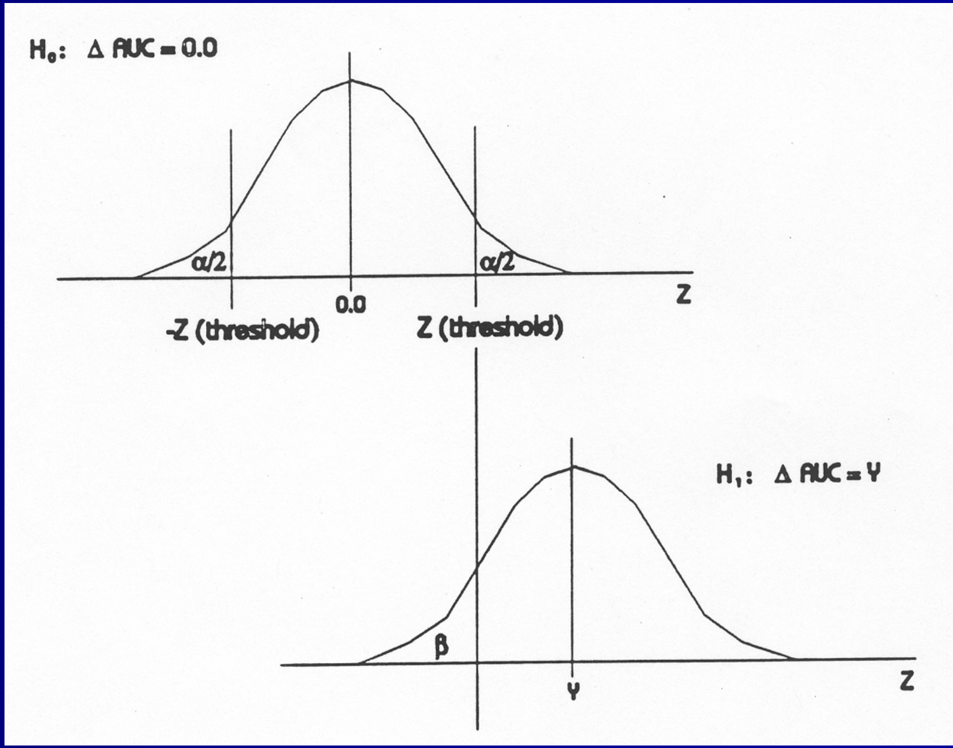
$$d = \left(\Theta_1 - \Theta_2\right)_n$$

N = number of observers

Evaluation of Observer Performance – analysis

Select α – the probability of incorrectly rejecting the null hypothesis.

For α and the error of the estimate, also report β – the probability of incorrectly failing to reject the null hypothesis when a specific alternate hypothesis is true.



Evaluation of Observer Performance – analysis

Widely-accepted software packages are freely distributed to calculate operating points, fit ROC curves, and perform statistical tests, including calculation of confidence intervals.

See http://xray.bsd.uchicago.edu/krl "ROC Analysis"

Well-respected researchers in ROC analysis include

Drs. Charles Metz, Kevin Berbaum, John Swets, Ronald Pickett, and Richard Swensson